Quarterly Journal of Experimental Psychology

ISSN: 0033-555X (Print) (Online) Journal homepage: http://www.tandfonline.com/loi/pqje19

Natural and contrived experience in a reasoning problem

P. C. Wason & Diana Shapiro

To cite this article: P. C. Wason & Diana Shapiro (1971) Natural and contrived experience in a reasoning problem, Quarterly Journal of Experimental Psychology, 23:1, 63-71, DOI: 10.1080/00335557143000068

To link to this article: http://dx.doi.org/10.1080/00335557143000068

Published online: 29 May 2007.

Quarterly Journal of Experimental Psychology (1971) **23**, 63–71

NATURAL AND CONTRIVED EXPERIENCE IN A REASONING PROBLEM

P. C. WASON AND DIANA SHAPIRO

Psycholinguistics Research Unit, Department of Phonetics, and Department of Psychology, University College London

This study is concerned with the effects of prior experience on a deceptive reasoning problem. In the first experiment the subjects (students) were presented with the problem after they had experienced its logical structure. This experience was, on the whole, ineffective in allowing subsequent insight to be gained into the problem. In the second experiment the problem was presented in "thematic" form to one group, and in abstract form to the other group. Ten out of 16 subjects solved it in the thematic group, as opposed to 2 out of 16 in the abstract group. Three hypotheses are proposed to account for this result.

Introduction

This study is about the effects of two kinds of experience on a deceptive reasoning problem. In the first experiment the experience is introduced as part of the procedure, and in the second it is inherent in the material used.

Previous experiments (Wason, 1968, 1969*a*) have established that it is very difficult to decide what information is required to test the truth of an abstract conditional sentence. For example, given the sentence: *Every card which has a D on one side has a* 3 *on the other side* (and knowledge that each card has a letter on one side and a number on the other side), together with four cards showing respectively D, K, 3, 7, hardly any individuals make the correct choice of cards to turn over (D and 7) in order to determine the truth of the sentence. This problem is called the "selection task" and the conditional sentence is called "the rule".

The rule has the logical form, "if p then q", where p refers to the stimulus mentioned in the antecedent (D); $\bar{p}$, i.e. not p, refers to the stimulus which negates it (K); q refers to the stimulus mentioned in the consequent (3); and $\bar{q}$, i.e. not q, refers to the stimulus which negates it (7). In order to solve the problem it is necessary and sufficient to choose p and $\bar{q}$, since if these stimuli were to occur on the same card the rule would be false but otherwise true.

The combined results of four experiments (see Table I) show that the subjects (students) are dominated by *verification* rather than *falsification*. On the whole, they failed to select $\bar{q}$, which could have falsified the rule, and they did select q, which could not have falsified it although this latter error is much less prevalent.

Experiment I

The previous experiments have been concerned with the stability of the errors and their resistance to correction by "remedial procedures". After the subjects had performed the selection task they had to evaluate the cards independently, i.e. turn them over and say whether the rule was true or false in relation to each. The present experiment is concerned with the prevention of error. The subjects are made familiar with the other side of the cards before the selection task is performed.

This prior experience is introduced by two methods. The "construction" method requires the subject to imagine, or project, a value on the other side of a card which would make the rule true, or make it false, in relation to it. In effect,

TABLE I

Frequency of the selection of cards in four previous experiments (n = 128)

p and q	59
p	42
p, q and $\bar{q}$	9
p and $\bar{q}$	5
others	13

positive and negative instances of the rule are constructed. The "evaluation" method simply requires the subject to turn over the card and say whether the rule is true, or false, in relation to it. The construction method clearly involves an imaginative act, and hence a greater degree of involvement than the evaluation method. It was accordingly predicted that it would be associated with superior performance in the subsequent selection task.

Design

Two independent groups were used: the construction group and the evaluation group. Both carried out their respective tasks on 24 cards in relation to a given conditional rule. They then performed the *initial selection task* with four more cards in relation to the *same* rule. A new conditional rule was then presented together with a further four cards. This *transfer selection task* was designed to assess the extent to which specific knowledge, gained in the prior experience, would be generalized.

Subjects

Twenty-four undergraduates (paid volunteers) of University College London were allocated alternately to the groups and tested individually. They had no previous experience with tasks of this type.

Procedure

Before presenting the rule all the subjects were first handed 28 cards, and instructed to inspect them to ensure that each had a letter of the alphabet on one side and a number on the other side.

They were then presented with the following rule: *Every card which has a vowel on one side has an even number on the other side.* Twenty-four of the 28 cards were then presented, one at a time, the remaining four being reserved for the transfer selection task. In the construction group they were instructed to name a value on the other side of each card which would make the rule true (or make it false). They were, however, told that it would be in order to say that no value on the other side would make the rule either true or false.

In the evaluation group they turned over each card and said that it made the rule true (or false). Similarly, they were told that it would be in order to say that a card was irrelevant to the truth or falsity of the rule.

The eight possible ways of permuting the logical values were each represented three times in the series of 24. They were presented successively in the following pairs, where the value given first refers to the symbol uppermost: (pq, $p\bar{q}$) ($\bar{p}\bar{q}$, $\bar{p}q$) (qp $q\bar{p}$) ($\bar{q}\bar{p}$, $\bar{q}p$). All the subjects received the cards in the same order, and within a pair the order of presenting the two cards was constant, but the pairs themselves were randomized in a different order within each of the three blocks of eight cards. In the construction group, where only the uppermost symbol was presented, the instruction for the first card within a pair was to name a symbol to make a verifying instance, and for the second card to name a symbol to make a falsifying instance.

In both groups the subjects were told they were wrong if they failed to evaluate (construct) $p\bar{q}$ and $\bar{q}p$ as falsifying, and if they did evaluate (construct) $\bar{p}q$ and $q\bar{p}$ as falsifying. This was to ensure that they did appreciate the falsifying instances of a conditional rule, but did not confuse them with the falsifying instances of an equivalence rule. The $\bar{p}q$ and $q\bar{p}$ instances do falsify an equivalence rule in the form: "if, and only if p then q".

For the initial selection task four cards (E, Z, 6, 7), taken from the 24 used in the prior experience, were placed on the table in a random order. The subjects were instructed that the rule now applied to these four cards taken as a whole, i.e. no longer independently. They were told "to select those cards, and only those cards, that would need to be turned over in order to discover whether the rule was true or false". No comments were made about these selections, and the subjects were not allowed to turn over any of the cards.

For the transfer selection task the following rule was presented: *Every card which has a D on one side has a* 3 *on the other side*, together with the four cards (D, K, 3, 5) which had not occurred in the series of 24, but had been included in the 28 originally inspected. The instructions were similar to those given for the initial selection task.

Results

Table II shows the frequency of correct and incorrect solutions, the first number in each cell referring to the initial selection task and the second to the transfer selection task.

TABLE II

Frequency of correct and incorrect solutions

	Correct	Incorrect	N
Construction	5 (6)	7 (6)	12
Evaluation	2 (2)	10 (10)	12
Totals	7 (8)	17 (16)	24

As predicted, there is a trend in favour of the construction group, but it falls short of statistical significance. The performance overall is unimpressive, particularly in the evaluation group. It will also be noted that the difference between the two selection tasks is negligible: knowledge is generalized to the extent that it has been gained. The two types of error, i.e. the selection of q and the omission of $\bar{q}$ are examined separately in Table III and IV.

Table III shows that both groups do better in omitting q than in getting the solution correct. But the frequency of this particular error also increases the difference between the groups in the predicted direction. On the transfer selection task it is significant ($P = 0{\cdot}05$, one-tailed, Fisher-Yates exact test).

It may be inferred from Table IV that the proportion of subjects in both groups who select $\bar{q}$ is greater than ever obtained initially in previous experiments. But it is also evident that none of the frequencies differ from chance expectancy. However, it may be inferred that the trend, showing the construction group superior on the correct solution, is entirely due to a greater tendency to omit q rather than one to select $\bar{q}$. The reasons for this result must be sought in the two different methods of introducing prior experience.

TABLE III

Frequency of selecting q

	q selected	q omitted	N
Construction	2 (3)	10 (9)	12
Evaluation	7 (8)	5 (4)	12
Totals	9 (11)	15 (13)	24

TABLE IV

Frequency of omitting $\bar{q}$

	$\bar{q}$ omitted	$\bar{q}$ selected	N
Construction	7 (5)	5 (7)	12
Evaluation	6 (6)	6 (6)	12
Totals	13 (11)	11 (13)	24

In the evaluation group the responses made in the prior experience corresponded to the "defective truth table" which is followed when a conditional sentence is evaluated (Johnson-Laird and Tagart, 1969). According to this truth table pq is classified as true, $p\bar{q}$ as false, and both $\bar{p}q$ and $\bar{p}\bar{q}$ as irrelevant. Only 1·4% of the evaluations deviated from this classification. In contrast, the truth table for the conditional in the propositional calculus counts all contingencies as true except $p\bar{q}$.

A very different picture emerges in the construction group. Table V shows the frequencies of instances constructed over the first block of eight trials, and (in parentheses) over the third block of trials.

Inspection of Table V suggests some tendency to reason by equivalence initially, i.e. to construct pq, qp, $\bar{p}\bar{q}$, $\bar{q}\bar{p}$, as verifying instances, and $p\bar{q}$, $\bar{q}p$, $\bar{p}q$, $q\bar{p}$, as falsifying instances. A third of the subjects were consistent in interpreting all the contingencies in this way over the first block of trials, and there was a very slight tendency for these subjects to perform better on the selection tasks. However, as a function of the feedback that $\bar{p}q$ and $q\bar{p}$ do not falsify, it will be observed that in the third block of trials there was a strong tendency to deny these contingencies falsifying status.

TABLE V

Frequencies of instances constructed on the first and third blocks of trials (n = 12)

Value presented		Value constructed				
		p	$\bar{p}$	q	$\bar{q}$	none
p	T?			12 (12)		0 (0)
	F?				12 (12)	0 (0)
$\bar{p}$	T?			1 (3)	7 (4)	4 (5)
	F?			6 (0)	1 (0)	5 (12)
q	T?	11 (10)				1 (2)
	F?		7 (1)			5 (11)
$\bar{q}$	T?		8 (7)			4 (5)
	F?	9 (10)	1 (0)			2 (2)

T = true, F = false. The numbers in parentheses refer to the third block of trials. The contingencies affected by feed-back are underlined.

This departure from the defective truth table may have been because the procedure was taken as a challenge to construct an instance. The subjects may have adopted a weaker standard of truth for the verifying instances, i.e. mere consistency with the rule. But when they proceeded to construct $\bar{p}q$ and $q\bar{p}$ as falsifying instances, they would have been corrected. The irrevelance of q would then have been learned directly, and this is reflected in the performance of the construction group in the selection tasks (see Table III).

It seems much more surprising that in both groups only about half the subjects selected $\bar{q}$ (see Table IV). An information-processing model, which has been devised to explain performance in these tasks (Johnson-Laird and Wason, 1970*a*), elucidates this result. The model postulates that the recognition that $\bar{q}$ could falsify the rule is not, in itself sufficient for its selection. Its relevance is assumed to depend on the arousal of a conflict between pq (as verifying) and $\bar{q}p$ (as falsifying). The conditions for this conflict occur if p had been selected and $\bar{q}$ omitted. But in the present experiment these conditions would be unlikely to have occurred because successive instances were constructed, or evaluated, independently of each other before the selection task is performed.

It may be concluded that the putative experience of logical structure, introduced procedurally, is relatively ineffective in enabling insight to be gained into the problem. It is reasonable to enquire whether "natural" experience, inherent in the subjects' everyday knowledge, may be more successful in inducing insight. It was predicted that when the material is realistic ("thematic"), as opposed to abstract, the selection task will be significantly easier.

Experiment II

Design

Two independent groups were used: the "thematic group" and the "abstract group" which differed solely in the terms in which the problem was presented.

Subjects

Thirty-two first year psychology undergraduates of University College London were allocated alternately to the groups and tested individually. They had no previous experience with tasks of this type.

Procedure

The thematic material represented a journey made on 4 different days of the week. Before presenting the rule about these journeys the subjects were given 16 cards which they inspected to ensure that each had the name of a town on one side and a mode of transport on the other side.

They were then presented with the four selection task cards, taken from the 16 originally presented, and arranged in random order on the table. They were instructed that they would now only be concerned with these cards. On two of them a different destination was written, i.e. "Manchester" and "Leeds", and on the other two a different mode of transport, i.e. "Car" and "Train". In addition each had a different day of the week in smaller type at the top.

The rule was then presented as a claim made by the experimenter about four journeys she had made on the four different days indicated on the cards. One variant of this rule was: *Every time I go to Manchester I travel by car.* Three other variants, derived from permuting the items on the cards, were also used. The presentation of all four was systematically rotated between the subjects to control for any possible preconceptions about the relation between destinations and modes of transport.

It was explained to the subjects that for each journey the destination appeared on one side of the card and the transport used on the other side. They were then instructed to say which cards they would need to turn over to decide whether the experimenter's claim was true or false. They were encouraged to take their time before answering.

A similar procedure was followed in the abstract group. Sixteen cards with a letter of the alphabet on one side and a number on the other side were first inspected. Four of these, D, K, 3, 7, were used for the selection task. The rule: *Every card which has a D on one side has a* 3 *on the other side,* was then presented as a claim made by the experimenter about the arrangement of letters and numbers on the cards. The subjects were instructed that this rule applied only to the four cards, and that they were to say which they would need to turn over to decide whether the claim was true or false.

Results

Table VI shows the frequency of correct and incorrect solutions.

TABLE VI

Frequency of correct and incorrect solutions

	Correct	Incorrect	*N*
Thematic	10	6	16
Abstract	2	14	16
Totals	12	20	32

The prediction that the thematic group would perform better than the abstract group is clearly confirmed by the distribution of the frequencies in Table VI ($P = 0{\cdot}004$, one-tailed, Fisher-Yates exact test). It is evident that representing the problem in the form of a realistic situation had a dramatic effect on the subjects' ability to gain insight into it. There may, however, be several reasons for this result.

Discussion

The results of the two experiments show the relative failure of procedurally introduced experience and the relative success of realistic material in allowing insight to be gained into the problem.

It could, of course, be argued that if the experience, introduced in Experiment I, had been more intensive, or if only the falsifying contingencies had been used, then performance would have been improved. But the purpose of the experience was only to acquaint the subjects with the logical structure of the problem, and not to train them to make particular responses. Previous results (Johnson-Laird and Wason, 1970*b*) have shown that various factors, such as cognitive load, may affect the appreciation of the task, and over-learning of the contingencies might be one more variable affecting performance. The point is that understanding the contingencies did not allow this knowledge to be used with maximum efficiency in the selection tasks. This result may seem incredible to anyone unacquainted with the difficulty of the problem. The reasons for it will not be discussed until the effects of thematic material on the task have been considered because these help to explain it.

Three hypotheses about different aspects of the thematic material used in Experiment II could account for its beneficial effects. First, the terms used in the thematic material, the towns and modes of transport, are concrete as opposed to the abstract terms which consisted of letters and numbers. It is well known that concrete material is better remembered than abstract material, and that in syllogistic reasoning familiar terms inhibit fallacious inferences (Wilkins, 1929). Thus in Experiment II the concrete terms may have been symbolically manipulated more readily and more appropriately than the abstract terms. This hypothesis might be tested by using concrete terms with an arbitrary connection, e.g. "Every card which has *iron* on one side has *apple* on the other side", where metals and fruits are known to occur on either side of the cards.

Second, it may be the concrete relation between the terms, rather than the terms themselves, which is beneficial. In the thematic material the relation which connects the terms is "travelling", as opposed to "the other side of the card" which connects the abstract material. This hypothesis could be tested by using abstract terms with a concrete relation between them, e.g. "Every time I go to K I travel by 3", where letters and numbers are known to stand for towns and transport respectively.

Third, the thematic material, unlike the abstract material, forms a coherent, unified whole: a claim about journeys supposed to have been made on four different days. Hence the subjects may have been more inclined to distribute their attention equally on its components, i.e. the four cards. They would thus be liberated from fixations on those cards which correspond to items mentioned in the rule. Cyril Burt (personal communication) has even suggested that thematic material enables the subjects to concentrate on the situation depicted, unfettered by the presence of the cards. This does not, in itself, explain why thematic material is helpful. But if it is assumed that knowledge about such material is represented in the brain in schemata, which may be activated by appropriate cues, then the solution to the problem may be simply "read off" by reference to this stored information.

The abstract material has no unifying link: each card is distinct and separate rather than being parts of a whole. The subjects are instructed that the rule refers only to the four cards, but in spite of this they may have construed it merely as a formula. They may, in fact, have regarded the cards as items in a sample from a larger universe, and reasoned about them inductively rather than deductively. In doing this they may have implicitly followed the Bayesian rule which assumes that the probability of a generalization is increased by repetition of confirming instances. Hence they might not have been disposed to consider the potential relevance of $\bar{q}$. There was some introspective support for probabilistic reasoning of this kind. It would follow, of course, that the experience of the problem's logical structure, introduced in Experiment I, would not have disabused the subjects of this particular misconception.

In fact, the difficulty of the abstract selection task may be due, not to the failure to recognize the correct solution, but to the failure to generate alternatives in order to derive the correct solution. In other words, abstract material may inhibit the realization of the necessity of combinatorial analysis rather than hindering the performance of such an analysis. The meaninglessness of the rule may tempt the subjects to interpret it, not as a rule, but as a sentence to be matched against instances. With thematic material it is gratuitous to talk about combinatorial analysis: the activation of stored knowledge spontaneously generates "real" alternatives. This hypothesis might be tested by comparing thematic and abstract material, but presenting all the possible solutions in a list from which one has to be selected, thus obviating the need for a combinatorial analysis. It would then be predicted only that the correct solution would be located more quickly with thematic material than with abstract material without a difference in its relative frequency.

Finally, the present results support the suggestion (Wason, 1969*b*) that it is not so much the logical structure which makes the abstract problem difficult, as the structure which the subjects impose upon the problem. Its difficulty does not lie in the fact that inferences of the kind demanded "hardly ever occur in real life"—a criticism sometimes voiced of the early experiments. On the contrary, when the task is made realistic it becomes appreciably easier. What makes the abstract task difficult is the arbitrariness of material which seems to defy the reasoning process. A more precise definition of the impediments involved must await further investigation.

The experiments in this paper form part of research to be reported in a thesis to be submitted for the degree of Ph.D. of London University by the second author, under the supervision of the first author. We are most indebted to our colleague, Dr P. N. Johnson-Laird, for invaluable critical comments and suggestions, and also to the Medical Research Council for a grant for scientific assistance.

References

Johnson-Laird, P. N. and Tagart, J. (1969). How implication is understood. *American Journal of Psychology*, **82**, 367–73.

Johnson-Laird, P. N. and Wason, P. C. (1970*a*). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology* **1**, 134–48.

Johnson-Laird, P. N. and Wason, P. C. (1970*b*). Insight into a logical relation. *Quarterly Journal of Experimental Psychology*, **22**, 49–61.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 273–81.

Wason, P. C. (1969*a*). Regression in reasoning? *British Journal of Psychology*, **60**, 471–80.

Wason, P. C. (1969*b*). Structural simplicity and psychological complexity. *Bulletin of the British Psychological Society*, **22**, 281–84.

Wilkins, M. C. (1929). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, No. 102.

Received 23 *April* 1970